doi: 10.14355/spr.2015.04.012

Classification of Familiar or Strange English Words using ERP

Yang Liu[1], Jiacai Zhang[*2]

College of Information Science and Technology, Beijing Normal University, Beijing, 100875, China

[1]ly20liuyang@163.com; [*2]jiacai.zhang@bnu.edu.cn

Abstract

English vocabulary is the bricks of the language, and it's important for English learners to handle accurately English words. Words recitation or dictation is the most widely used simple test form in vocabulary quiz. However, some students do less well in tests due to stress or even meltdown for exam. Here, we apply Brain fingerprinting (BF) in English vocabulary test. We studied the Event Related Potential (ERP) components elicited by familiar English words and strange English words, and used the ERP to decipher whether subject recognized the word presented in the screen. We analyzed the distinct ERP features related to different stimuli (familiar and strange words), and further speculated whether subjects recognized the presenting words from P300 by calculation correlations between different stimuli. The results of 8 subjects implied that familiar words recognized by subjects could be detected from P300 with accepted precision and recall. This means that BF provides a novel alternative for vocabulary quiz and opens the way for application in education, where concealed vocabulary related information stored in mind was detected by measuring brain signals, such as electroencephalography (EEG).

Keywords

Brain Fingerprinting; P300; Event-related Potential; Vocabulary Quiz

Introduction

The strength of students' vocabulary is very important to evaluate student English level. The more words students are able to handle accurately, the better their chances of understanding English and making themselves understood. Words recitation and dictation, as the simple vocabulary test, is most used in English exam.

However, explicit vocabulary exams always bring great pressure to students. Some students even feel excessively worry about the upcoming exam and fear to be evaluated. Too much exam stress can cause individuals to perform poorly on tests. Here, we employed Brain fingerprinting (BF) technique to evaluate students' vocabulary implicitly.

BF is a technique that uses electroencephalography (EEG) to determine whether subject's brain stored specific information related to known or relevant item by measuring electrical brainwaves[1,2,3]. EEG Signal was recorded when pictures or words was presented. A specific signal would be evoked in brain activity if the pictures or words coincide with the memory of the subject. And this specific signal can be detected in electroencephalographic (EEG) brainwaves, i.e. P300, a specific EEG event-related potential. So detection of P300 reflects whether known or relevant information exists. In the previous literature, the P300-based detection has been well used in guilty knowledge tests [3,8]. Then laboratory and filed tested at the FBI, CIA, and US Navy have got good results with 0% errors [1,2]. In studies above, subjects usually had two roles, guilty or innocent, and researchers deciphered the role from the brain responses. Where, three kinds of stimuli were employed: probe (P), irrelevant (I), and target (T). The decision of subjects' role is based on the fact that the brain's processing of known information, such as the details of a crime stored in the brain, is revealed by a specific potential in P300 response.

In this study, we are aimed to decipher whether subjects recognize the stimuli words, rather than the subjects' role. We classified familiar words recognized by students from unacquainted words using EEG brain responses recorded during subjects viewed the words on the screen. Experimental results demonstrated high classification accuracy for all 8 subjects, showed that distinct P300 features were elicited by memory-relevant words, and this paved the novel application of BF in education, such as vocabulary quiz.

Materials and Methods

Dataset

10 student subjects (5 males, 5 females) from Beijing Normal University (BNU) participated in the experiment. The native language of all subjects is Chinese. Prior to the experiment, all subjects were provided with written informed consents to participate.

There were three kinds of stimuli presented to subjects:

- Probe (P): memory-relevant words that have been recently learned. The probe words were non-existent words, half of the probes were presented to subjects before the experiments and subjects were asked to remember them. Some examples of P: chickic, pumiper, vegeat.
- Irrelevant (I): new words that were memory-irrelevant. The irrelevant words were made up, and they are non-existent. Some examples of I: idanabe、drgeaio、clandibze.
- Target (T): study-relevant words that were previously learned. The materials for stimuli T were commonly-used words. Some example of T: turtle、whale、elephant.

The stimuli P and I were some fabricated words whose lengths were similar to that of T. Prior to the experiment, subjects got a paper with half of all the probe words and target words on it, and were asked to memory them. The half of the probes was learned or known words, and the rest of words were strange or unknown words.

Subjects were tested in three sessions, and each session contained 144 trials. On each trial we presented the subject with a stimulus word on the screen. The words from T, P, or I were presented in a random order. In each session, 6 target items, 6 probe items, and 24 irrelevant items, totally 36 stimuli were repeated for 4 times, which was shown in Table 1. Each trial consisted of stimulus display of 0.3 second and an inter-stimulus interval of 1.8 seconds. We had a fixation point of 5 seconds for relaxation between every 36-stimuli presentation.

TABLE 1 STIMULI

Stimulus Type	Probability	Description	Task
Target	1/6	Irrelevant to study	Button press
Irrelevant	2/3	Irrelevant to study	No button press
Probe	1/6	Memory-relevant words and relevant to study	No button press

We also recorded subjects' behavior data. Subjects were asked to push a button only when they saw the target word appeared on the screen. Only the data of those subjects who marked out more than 90% of the target words were used for later analysis. Two male subjects were removed due to more than 10% error in behavior data. Finally, data from 8 subjects (3 male, 5 female) were chosen to be used in the following analysis.

EEG data were collected from frontal and parietal area by G.HIamp record system. The sample frequency is 256 Hz. And the EEG data were then performed band-pass filter[0.1-30 Hz].

Data Analysis

10 student subjects (5 males, 5 females, and aged 20-24) from Beijing Normal University (BNU) participated in the experiment. Prior to the experiment, all subjects were provided with written informed consents to participate. The purpose of data analysis is to retrieve the probe words by comparing the correlation between the response to probe with the average response to target words, and the correlation between the response to probe with the average response to irrelevant words. The probe word was determined as acquainted words only if the P-T correlation exceeds the P-I correlation. In other words, If the probe words responses were more similar to the responses to target words, the subject was determined to be "information present", or he/she had known the probe words (acquaintance). If the responses to probe words were more similar to the responses to irrelevant words, the subject was determined to be "information absent", or he/she had not known the probe words(stranger).

Our purpose is to assess the correlation between the probe ERPs and target or irrelevant ERPs. We want to

determine whether subjects know the probe words by comparing the correlation. Prior to the waveform analysis, all waveforms were digitally band-pass filtered(1-6Hz) [1].We averaged the T data and I data to get an average T waveform and an I waveform respectively, and then computed the P-T correlation and P-I correlation. With the threshold value, we determined words to be strange or familiar.

Result

The average ERPs for target, probe and irrelevant responses were displayed in Fig. 1. As to be seen in the figures, the amplitude elicited by target was larger than that of irrelevant words. It's notable that the waveform for learned probe was close to that of target, while the waveform for strange probe was similar to that of irrelevant words.

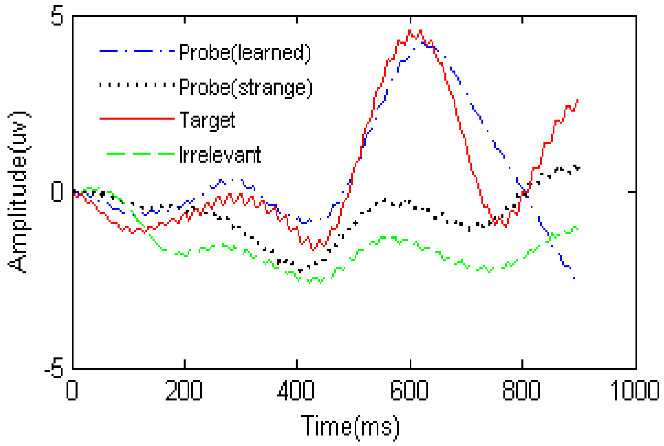

FIG. 1 AVERAGE ERPS FOR DIFFERENT WORDS

For all 8 subjects' results, the accuracy to classify learned and strange probes was shown in Fig. 2, and the mean accuracy was 76% with standard deviation 9.4%. The best subjects got accuracy of 91.7%, and all subjects got accuracy far above chance level.

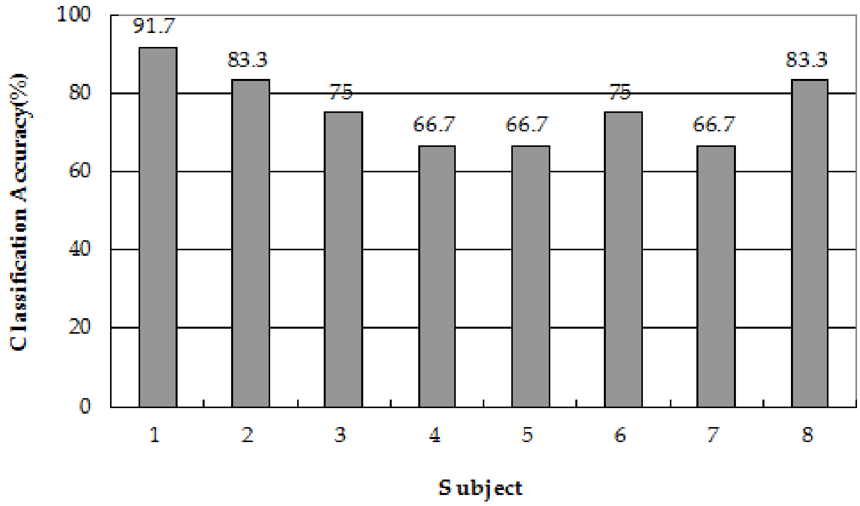

FIG. 2 CLASSIFICATION ACCURACY FOR ALL SUBJECTS

Conclusion

The experiment results in this study confirm the availability of the P300-based detection for memory words applications. The current study requires to be validated with more subjects. Improvement of the probe classification accuracy is also necessary for the further research. However, the current study offers a step forward in the development of the P300-based vocabulary quiz and other memory detections related to application in education.

ACKNOWLEDGMENT

This work is supported by 863 Program (2012AA011603). This work is also supported by the NSFC Key Program (91320201), and General Program (61375116).

REFERENCES

[1] Farwell L.A.,"Brain fingerprinting: a comprehensive tutorial review of detection of concealed information with event-related brain potentials," Cogn Neurodyn 6 (2012): 115-154.

[2] Farwell L.A., Richardson D.C., Richardson G.M.,"Brain fingerprinting field studies comparing P300-MERMER and P300 brainwave responses in the detection of concealed information," Cogn Neurodyn 7 (2013): 263-299.

[3] Farwell L.A., Donchin E.,"The truth will out: interrogative polygraphy ("lie detection") with event-related potentials," Psychophysiology 28(5) (1991): 531-547.

[4] Farwell L.A., Donchi E.,"The "brain detection": P300 in the detection of deception," Psychophysiology 23(4) (1986): 434

[5] Rosenfeld J.P., Biroschak J.P., Furedy J.J.,"P300-based detection of concealed autobiographical versus incidentally acquired information in target and non-target paradigms," International Journal of Psychophysiology 60 (2006): 251-259.

[6] Rosenfeld J.P., Soskins M., Bosh G., Ryan A.,"Simple, effective countermeasures to P300-based tests of detection of concealed information," Psychophysiology 41 (2004): 205-219.

[7] Allen J.J., Iacono W.G.,"A comparison of method for the analysis of event-related potential in deception detection," Psychophysiology 34 (1997): 234-240.

[8] Abootalebi V., Moradi M.H., Khalizadeh M.A.,"A comparison of methods for ERP assessment in a P300-based GKT," International Journal of Psychophysiology 62 (2006): 309-320.

[9] Miyake Y., Mizutanti M., Yamahura T.,"Event related potentials as an indicator of detection information in field polygraph examinations," Polygraph 22 (1993): 131-149.

[10] Abootalebi V., Moradi M.H., Khalizadeh M.A.,"A new approach for EEG feature extraction in P300-based lie detection.," Computer Methods and Programs in Biomedicine 94 (2009): 48-57.

[11] Johnson M.M., Rosenfeld J.P.,"Oddball-evoked P300-based method of deception detection in the laboratory. Ⅱ: utilization of non-selective activation of relevant knowledge," Int J Psychophysiol 12(3) (1992): 289-306.

[12] Johnson R.J.,"On the neural generators of the P300 component of the event-related potential," Psychophysiology 30(1) (1993): 90-97.